

# Reference Dependence Effects on WTA and WTP Value Functions and Their Disparity

Mark J. Koetse<sup>1</sup>  · Roy Brouwer<sup>1</sup>

Accepted: 15 April 2015 / Published online: 12 May 2015  
© The Author(s) 2015. This article is published with open access at Springerlink.com

**Abstract** In this study we investigate two reference dependence effects in a choice experiment. The first is the effect of the well-known distinction between gains and losses, the second is the effect of changing the reference value on willingness to pay (WTP) and willingness to accept (WTA). The latter has to our knowledge not been studied before. We hypothesize that there are differences between WTA and WTP, and that both value functions and their disparity are affected by the absolute value of the reference point. We test our hypotheses using a choice experiment with trade-offs between changes in flood probabilities and costs. The choice experiments elicit WTP and WTA, using two flood probability reference values, yielding four separate value functions. Our findings show that a substantial WTA–WTP disparity exists, and that this disparity increases when moving away from the reference point. Also, both WTA and WTP value functions are affected by the flood probability reference value, and the WTA–WTP disparity increases when the flood probability reference point increases. Both findings suggest that welfare effects caused by changes in public good provision depend not only on the direction of change (*loss aversion*), but also the *reference value*. Moreover, our results show that the latter effect is larger for losses than for gains. We introduce the concept of reference point updating as a possible explanation for these findings.

**Keywords** Value functions · WTA–WTP disparity · Prospect theory · Reference dependence · Choice experiment · Flood probability · Reference point updating

**JEL Classification** D80 · H40 · Q25 · Q54

---

✉ Mark J. Koetse  
mark.koetse@vu.nl

Roy Brouwer  
roy.brouwer@vu.nl

<sup>1</sup> Department of Environmental Economics, Institute for Environmental Studies, VU University Amsterdam, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands

## 1 Introduction

The pervasive disparity between willingness to accept (WTA) and willingness to pay (WTP) for identical changes in goods and services implies that choosing appropriate welfare measures to assess the economic consequences of those changes is essential. This is especially true in the realm of public goods provision, where property rights are not always well defined and market prices to guide welfare analysis and decision making are often missing. Substantial differences between WTA and WTP were first observed in hypothetical questions involving public goods (see [Cummings et al. 1986](#)). Since then many empirical studies have, almost without exception, reported large WTA–WTP disparities (e.g., [Horowitz and McConnell 2002](#); [List 2003](#)). Within an expected utility framework, substitution and income effects have been put forward to explain the observed WTA–WTP disparities (e.g., [Hanemann 1991](#)). However, studies by [Sudgen \(1999\)](#) and [Horowitz and McConnell \(2003\)](#), among others, show that the observed differences cannot be reasonably explained by income and substitution effects only.

With expected utility theory failing to fully explain the patterns observed in the empirical literature, prospect theory has become the main alternative theory of decision behaviour under risk since [Kahneman and Tversky \(1979\)](#). Key elements of the original version of the theory are that: (1) people value losses higher than gains, leading to a disparity between WTA and WTP (loss aversion); (2) the marginal value of both gains and losses decreases with their magnitude. These two properties give rise to an S-shaped value function that is asymmetric around the reference point (steeper for losses than for gains). Other, extended versions of prospect theory have been developed since [Kahneman and Tversky \(1979\)](#). For example, cumulative prospect theory considers the cumulative probability distribution function of all prospects instead of the individual prospect probabilities (also known as *rank dependence*), and allows for different probability weighting functions of gains and losses ([Tversky and Kahneman 1992](#)). Various other theoretical models of reference dependence, for situations with and without risk, have been developed in recent studies (e.g., [Kőszegi and Rabin 2006](#); [Neilson 2006](#); [Loomes et al. 2009](#)).

Empirical evidence on loss aversion is abundant in general (e.g., [Horowitz and McConnell 2002](#)), but for preferences regarding public environmental good provision it is limited (for exceptions, see [Bateman et al. 2009](#), and [Brouwer and Schaafsma 2013](#)). In this paper we add to the literature by empirically testing the combined effects of loss aversion and changing the value of the reference point on WTA and WTP value functions and the WTA–WTP disparity. To our knowledge these latter effects have not been studied before. For this purpose we develop a choice experiment in a case study area where changes in future flood probabilities may cause substantial damage. Stated preference methods such as contingent valuation and choice experiments are rarely used in flood valuation studies ([Brouwer et al. 2009](#); [Dekker et al. 2014](#)), but there is an emerging stated preference literature on demand for flood insurance, both in developed and developing countries ([Botzen and Van den Bergh 2009, 2012](#); [Brouwer and Akter 2010](#); [Brouwer et al. 2014](#)). These studies typically investigate household WTP for flood insurance under different flood probabilities and damage cover. The general expectation underlying WTP for flood control or insurance is that people are risk averse when their decision involves potential losses under low probability-high impact conditions, and the corresponding individual choice behaviour is motivated by a desire for security. A study by [Brouwer and Schaafsma \(2013\)](#) is the only study we know of that examines both WTP and WTA in the context of increasing flood probabilities.

The remainder of this paper is organised as follows. Section 2 presents the theoretical background and our research hypotheses. Section 3 discusses the design of the choice experiments and the welfare measures used. Section 4 is dedicated to statistical design and data collection, while Sect. 5 discusses results of statistical tests and the statistical model used for data analysis. In Sect. 6 we discuss the estimation results. Section 7 is aimed at hypothesis testing and presents our findings regarding the effects of reference dependence on the WTA and WTP value functions, and on the WTA–WTP disparity. Section 8 concludes.

## 2 Theoretical Background and Hypotheses

Our starting point is a model that was recently used in an experimental study on reference dependence by [Viscusi and Huber \(2012\)](#). Their model is originally based on the model developed by [Kőszegi and Rabin \(2006\)](#), in which the utility impact of a change in the consumption of a good or service depends on the difference between the new level and a reference level (like [Tversky and Kahneman 1991](#)), and on the absolute value of the new level of consumption (unlike [Tversky and Kahneman 1991](#)). This model is of specific interest for our purposes because it is used to study trade-offs between changes in costs and risk probabilities. We also study people's trade-offs between costs and risk probabilities, but while [Viscusi and Huber \(2012\)](#) focus on separating the effects of cost and health probability reference dependence, we focus on the combined impact of loss aversion and changing the flood probability reference value on the WTA and WTP value functions and the WTA–WTP disparity.

[Viscusi and Huber \(2012\)](#) consider a situation in which a person goes from an initial reference situation  $r$  with monetary cost  $c_r$  and health risk probability  $p_r$ , to a new situation  $n$  with monetary cost  $c_n$  and health risk probability  $p_n$ . For our purposes, we change health risk to flood risk, but otherwise our set-up is very similar. Using a slightly different notation than in [Viscusi and Huber \(2012\)](#), utility ( $U$ ) in the new situation is given by:

$$U(c_n|c_r; p_n, k|p_r, k) = v(c_n) + \mu[v(c_n) - v(c_r)] + p_n h(k) + \lambda[p_n - p_r]h(k), \quad (1)$$

where  $v(c)$  is the utility function of monetary costs (for flood control measures),  $k$  represents the damage costs associated with a flood,  $h(k)$  is the flood (damage) cost utility function, and  $\mu$  and  $\lambda$  are reference dependence parameters for monetary costs and flood risk probability, respectively. In this study we keep the individual damage costs related to floods fixed, and introduce a trade-off between the monetary costs of flood prevention and flood probability. As usual in reference dependence models, both  $\mu$  and  $\lambda$  are larger for losses than for gains ( $\mu^- > \mu^+$  and  $\lambda^- > \lambda^+$ ). [Viscusi and Huber \(2012\)](#) focus on separating the effects of cost reference dependence and health probability reference dependence, and find that the former impacts utility more than the latter. Moreover, their results show that health probability reference dependence only matters in the domain of cost decreases. Although our set-up is similar, we focus on the combined effects of flood probability loss aversion and changing the absolute value of the flood probability reference value. With respect to the latter we introduce a concept we call 'reference point updating'. This may occur when a change occurs stepwise during several stages (e.g., changes through time). The underlying intuition is that once a change has occurred, people adjust to the new situation and update their utility function to account for the change in reference point. If another change occurs, people value that change using their new reference point instead of the initial one, and consequently the value

attached to the change is different than if it would have occurred simultaneously with the first change.<sup>1</sup>

In our model, a change in utility  $dU/dp$  caused by a change in flood probability from  $p_r$  to  $p_n$  is given by:

$$dU(p_n|p_r) = -(1 + \lambda)(p_n - p_r)h(k). \quad (2)$$

Now consider a decrease in flood probability from  $p_1$  to  $p_3$ . Reference point updating would mean that going from  $p_1$  to  $p_3$  produces a different change in utility as going first from  $p_1$  to  $p_2$ , with  $p_1 > p_2 > p_3$ , and then from  $p_2$  to  $p_3$ . In mathematical terms, using Eq. (2), this boils down to the following inequality:

$$-(1 + \lambda)(p_3 - p_1)h(k) \neq -(1 + \lambda)(p_2 - p_1)h(k) - (1 + \lambda)(p_3 - p_2)h(k). \quad (3)$$

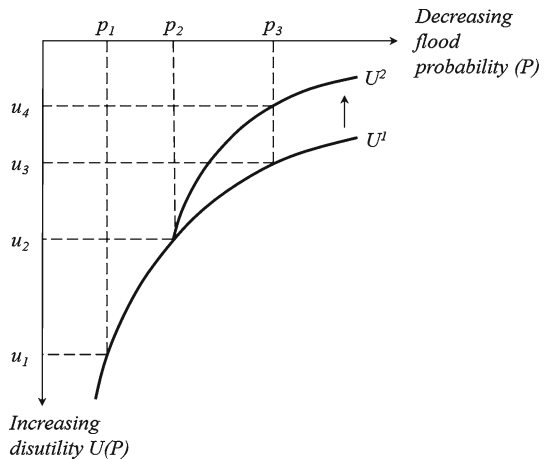
If  $\lambda$  is independent from  $p$ , the terms on the right hand side can be rewritten to yield the term on the left hand side, implying that the inequality only holds when  $\lambda$  is a function of the flood probability given its reference point, i.e.,  $\lambda = f(p_n|p_r)$ . With respect to our initial model formulated in Eq. (1), this inequality would imply that the model is incomplete. An updated version would need to incorporate some form of  $\lambda = f(p_n|p_r)$ , the purpose of which would be to incorporate a reference point updating process.

We present this process graphically in Fig. 1. In the figure the  $x$ -axis represents flood probability, which decreases when moving to the right, and the  $y$ -axis represents the disutility related to flood probability. Utility increases (or rather, disutility decreases) when flood probability decreases ( $U'(p) < 0$ ) and we assume (and will later on show) that marginal utility increases when flood probability increases ( $U''(p) > 0$ ). The figure represents a situation in which a person is confronted with flood probability  $p_1$ , with negative utility  $u_1$  derived from utility curve  $U^1$ . An exogenous change now occurs which causes flood probability to move from  $p_1$  to  $p_3$ , leading to a change in the associated utility of  $(u_3 - u_1)$ . Now consider a two-step decrease in flood probability, first from  $p_1$  to  $p_2$  and then from  $p_2$  to  $p_3$ . When there is no reference point updating, i.e., the utility curve does not shift, these changes lead to an identical change in utility as before ( $u_3 - u_1$ ). However, under reference point updating, the flood probability utility curve is expected to shift upwards to  $U^2$ , and the two-step change in flood probability induces a change in utility to  $u_4$ . This implies a larger change in utility than under no updating, i.e.,  $(u_4 - u_1) > (u_3 - u_1)$ . The main idea here is that changing the reference value in a choice situation affects the location of the utility curve and thereby the WTP and WTA value functions for preventing or accepting a change, in our case a change in flood probability.

To investigate the combined impact of loss aversion and changing the absolute value of the reference point, we design our study such that we are able to derive value functions using separate flood probability reference values for both losses and gains, i.e.,  $U^1(p_{r1})$  and  $U^2(p_{r2})$  with  $p_{r1} > p_{r2}$ . As in Eq. (1), we define  $\mu^+$  and  $\mu^-$  for respectively cost decreases (gains) and cost increases (losses), and  $\lambda^+$  for respectively flood probability decreases (gains) and  $\lambda^-$  for probability increases (losses). In our design we do not change the reference value for costs, although we distinguish between cost increases and decreases in order to be able to measure both WTA and WTP. We estimate WTA and WTP on identical parts of the utility curve, implying that the flood probability reference point for WTP ( $p_1$  in Fig. 1) is opposite the one for WTA ( $p_3$  in Fig. 1). This means that reference point updating for WTP should

<sup>1</sup> Although reference point updating may seem to be related to Bayesian updating, the two processes are different. In our case there is no individual updating due to new information, which is an essential element of Bayesian updating (see, e.g., Greenberg 2008). We compare valuation results across split samples of homogeneous respondents who were given different reference starting points.

**Fig. 1** Graphical representation of reference point updating: the flood probability utility function  $U(P)$  is updated from  $U^1$  to  $U^2$  due to a decrease in the flood probability ( $P$ ) reference value from  $p_1$  to  $p_2$



occur when the flood probability reference point decreases (e.g., from  $p_1$  to  $p_2$  in Fig. 1), and for WTA when the flood probability reference point increases (e.g., from  $p_3$  to  $p_2$  in Fig. 1). Based on the considerations discussed above, we test the following two hypotheses:

Hypothesis 1: There is a disparity between WTA and WTP for identical flood probability changes:  $\lambda^+ < \lambda^-$ .

Hypothesis 2: WTA and WTP flood probability value functions move upward as the flood probability reference value decreases for WTP and increases for WTA.

To our knowledge this is the first study that addresses the impact of varying reference points on the WTP–WTA disparity. There are no existing empirical insights into the effects of different reference points on WTA and WTP value functions. Since we measure WTP and WTA on identical parts of the utility curve, we introduce opposite flood probability reference points for WTA and WTP. In this situation, and as defined in hypothesis 2, updating is expected to take place when the flood probability reference value decreases for WTP and increases for WTA. This implies an increase in WTP and a decrease in WTA when the flood probability reference value decreases, leading to an unambiguous decrease in the WTA–WTP disparity. This leads to our third hypothesis:

Hypothesis 3: The disparity between WTA and WTP decreases as the flood probability reference value decreases.

In order to test our hypotheses, we carry out choice experiments to measure consumer preferences in the specific context of flood risks. In four separate experiments, identically sampled groups of consumers exposed to the same flood risk are asked to value changes in flood probability, starting from different flood probability reference points and differently defined property rights (WTP or WTA). One group is given the right to a certain risk protection level (WTA), whereas the other is not (WTP). This allows us to test possible differences between WTP for avoiding and WTA compensation for increased flood probabilities. Incorporating different flood probability reference points allows us furthermore to test the possible existence of reference point updating. The four experiments and their interrelations are summarised in Table 1.

**Table 1** The four experiments and their interrelations

	Probability reference value = $p_{r1}$	Probability reference value = $p_{r2}$
Willingness to pay (WTP) to prevent an increase in flood probability	WTP1 $-(1 + \lambda^+) (p_n - p_{r1}) h(k)$	WTP2 $-(1 + \lambda^+) (p_n - p_{r2}) h(k)$
Willingness to accept (WTA) compensation for an increase in flood probability	WTA1 $-(1 + \lambda^-) (p_n - p_{r1}) h(k)$	WTA2 $-(1 + \lambda^-) (p_n - p_{r2}) h(k)$

By comparing results between WTP1 and WTA1, and between WTP2 and WTA2, we test the disparity between WTA and WTP. By deriving and comparing value functions for WTP1 and WTP2, and for WTA1 and WTA2, we test our second and third hypothesis, in particular to what extent the disparity between WTP1 and WTA1 on the one hand, and the disparity between WTP2 and WTA2 on the other, changes as a result of varying flood probability reference points.

### 3 Choice Experiment Design and Welfare Measures

The study is carried out around the IJsselmeer, one of the largest freshwater buffers in Europe (Fig. 2). This buffer is used during the summer as one of the main sources of water supply for agriculture and residential household water demand. In the winter the lake functions as a buffer for excess storm and river water. In order to anticipate future climate change and corresponding droughts and floods, the Dutch government is considering a future increase in the IJsselmeer water level.

Increases in the water level have a number of consequences in and around the IJsselmeer. Without additional government investments in flood control, flood probabilities will increase, wildlife habitat along the shores of the IJsselmeer will disappear, and bird populations will decrease in size and variation. In this study we assess preferences of people who live around the IJsselmeer for these potential future changes. In our experiments we include one label, i.e., raising the dikes, and four attributes, i.e., creating shores along the coast of the IJsselmeer, changes in the bird population, changes in flood probability, and changes in the annual municipal tax.

In view of the fact that increases in the height of dikes are correlated with flood probabilities, raising the dikes around the IJsselmeer is included as a label, indicating whether dike increases take place or not. With this we aim to measure the influence of increasing the height of dikes on safety perception, whilst controlling for changes in flood probability.

Economically optimal flood probabilities around the IJsselmeer vary from once every 10,000 years to once every 500 years (Kind 2014), so we use this range in our experiment. It is unlikely that respondents know current flood probabilities, not least because they differ per location (Bos et al. 2012). Using respondent-specific levels is therefore not sensible. In order to be able to test our hypotheses, we design a WTP experiment and a WTA experiment, each with two different reference situations, resulting in four separate experiments: WTP1, WTA1, WTP2, and WTA2. In each experiment the absolute flood probability levels vary but



**Fig. 2** Location of the IJsselmeer study area (*blue*) and sample population around the IJsselmeer (*red*) in the Netherlands. (Color figure online)

the relative change in flood probabilities is kept constant. We furthermore ensure that the ranges of the levels are largely overlapping. The levels in the first set of WTP and WTA experiments (WTP1 and WTA1) are once every 5,000 years, once every 1,000 years and once every 500 years. The levels in the second set of WTA and WTP experiments (WTP2 and WTA2) are once every 10,000 years, once every 2,000 years and once every 1,000 years. In both experiments respondents were informed that in the future these are the levels they could be facing due to increasing water levels. Two pre-test rounds were used to assess how well respondents understood and could meaningfully distinguish between the attribute levels. We used a risk ladder so respondents could assess flood risks compared to a variety of other risky events. Our pre-tests also showed that presenting flood probabilities as the chance that one or more floods occur during a 30-year period, the standard mortgage time period in the Netherlands, considerably improved understanding of the presented probabilities. Hence, flood probabilities were presented to respondents in both ways. For example, a flood probability of once every 10,000 years is equivalent to a probability of 0.3 % of at least one flood in 30 years, while a flood probability of once every 500 years is equivalent to a probability of 6 % of at least one flood in 30 years.

An alternative means of flood control is the creation of shores to break the waves coming from the IJsselmeer. The shores also provide a wildlife habitat shelter and have positive effects on nature and landscape aesthetics. Also the costs of building and maintaining dikes and shores differ substantially, and therefore there exists a trade-off between public safety, nature and costs. We distinguish between three levels: no shores, shores off the coast and shores next to the dike.



**Table 2** Attributes and attribute levels in the four choice experiments

Attributes	Attribute levels
Raising the dikes	No (reference scenario in WTP version) Yes (reference scenario in WTA version)
Flood probability	Once every 5,000 years (reference scenario WTA1)
WTP1 & WTA1	Once every 1,000 years Once every 500 years (reference scenario WTP1)
Flood probability	Once every 10,000 years (reference scenario WTA2)
WTP2 & WTA2	Once every 2,000 years Once every 1,000 years (reference scenario WTP2)
Change in bird population	0 % (reference scenario WTA) –10 % –30 % (reference scenario WTP)
Type of shore	No shores (reference scenario WTP) Shores off the coast Shores next to the dike (reference scenario WTA)
Change in annual local tax (increases for WTP, decreases for WTA)	€0 (reference scenario) €60 €100 €180

Increasing the IJsselmeer water level is expected to have substantial effects on the bird population in the area as it changes the composition of the existing natural areas in the IJsselmeer area. Increasing water levels is expected to reduce bird populations by as much as 30 %. Following expert judgment (Bos et al. 2012), we include bird population reductions of 0, 10 and 30 % in the experiment.

The costs of raising dikes and building shores will be paid for through local taxes. The tax levels need to be sufficiently wide to cover maximum WTP and minimum WTA. After varying and testing the levels in two pre-test rounds, changes in annual local tax (increases for WTP and decreases for WTA) were included of 60, 100 and 180 Euro, while there is no change in annual local tax in the reference scenario. The attributes and attribute levels for the four experiments are summarised in Table 2.

We are interested in the WTP and WTA for identical parts of the utility curve, so we use identical attribute levels, but switch the reference points in the WTP and WTA versions of our choice experiment. Nevertheless, the WTP for avoiding a negative environmental change thereby measures an *equivalent surplus*, while the WTA a negative environmental change measures a *compensating surplus*. Although this difference may to some extent affect the difference between WTP and WTA, it does not affect the effects of changing the reference value (i.e., effects caused by changing the flood probability reference points in experiments 1 and 2).

Another aspect of the design is related to cost reference dependence. Viscusi and Huber (2012) find that reference dependence with respect to costs is stronger than health probability reference dependence. In our design the effects of changing the flood probability reference value on the WTA and WTP estimates, and on the WTA–WTP disparity, cannot be explained by costs reference dependence, since the experiments with different flood probability refer-



ence points use identical cost reference points and identical cost changes, both in terms of direction and magnitude, even though the WTP questions involve a cost *increase*, while the WTA questions involve a cost *decrease*.

## 4 Statistical Design and Data Collection

The attribute levels can be combined to generate 81 alternative policy scenarios. These cannot all be shown to respondents, hence a fractional factorial main effects design was generated using the software Sawtooth CBC. Sawtooth uses a randomized design strategy, and produces a design that is close to orthogonal with minimal overlap of attribute levels and efficient in terms of main effects. During statistical design generation we imposed the restriction that the policy scenarios should include an improvement (in the WTP version) or deterioration (in the WTA version) of at least one of the three non-monetary attributes. The statistical design consists of 15 versions of 10 choice tasks each, and each respondent was randomly assigned to one of these 15 versions. In order to ensure an identical design and identical trade-offs for the WTP and WTA versions, the attribute levels for local tax were switched around in each choice question for the WTA version compared to the WTP version.<sup>2</sup>



The reference situation in the WTP experiment is a situation that is most unfavourable with respect to flood probability, presence of shores and bird populations, but in which no tax increases take place. The two alternative policy scenarios in the WTP version are more favourable with respect to at least one of these three non-monetary attributes, accompanied by an increase in annual local taxes. The reference situation in the WTA version is a situation that is most favourable with respect to flood probability, shores and bird populations, but in which there are no tax reductions. The two alternative policy scenarios in the WTA version are less favourable with respect to at least one of the three non-monetary attributes, and are accompanied by a compensation in the form of a decrease in annual local taxes. Examples of WTP and WTA choice tasks are presented in Fig. 3a, b. Each respondent was first shown an example of a choice card with instructions about the choice task. After that respondents answered ten choice tasks, followed by debriefing questions about the way in which they made their choices.<sup>3</sup>

The survey was implemented via internet in December 2012. The relevant population in this study consists of people who are expected to be directly affected by potential future changes in the IJsselmeer. The sample was drawn from a Dutch internet panel owned by TNS-NIPO, which includes more than 200,000 households. The panel is established through random sampling, meaning that each household has an equal chance to be added to the panel as long as it conveyed a willingness to participate. In the TNS-NIPO panel around 6,800 respondents live in the postal code areas around the IJsselmeer (red areas in Fig. 1). From this set, four independent samples were drawn, one for each of the four choice experiment versions. For each sample we employed representative sampling based on age, gender, household size, social class (education and profession) and residential location. For each of the four choice experiments 375 households were invited to complete the survey. The







<sup>2</sup> For example, an increase of 180 Euro in the WTP version in choice option 1 vis-à-vis an increase of 100 Euro in choice option 2 implies that choice option 2 is financially more attractive by 80 Euro. Now suppose we use the same design for the WTA version, and note that tax increases in the WTP version have become tax decreases in the WTA version. In this case choice option 1 has become financially more attractive than choice option 2 by 80 Euro. Therefore, in order to make the designs for WTA and WTP identical, the levels for local tax have to be switched around for the WTA version compared to the WTP version.

<sup>3</sup> The full survey is available upon request from the authors.

(a)

	OPTION 1 No increase in dike height	OPTION 2 Increase in dike height	OPTION 3 Increase in dike height
Flood probability	Once every 1,000 years (3% in 30 years)	Once every 1,000 years (3% in 30 years)	Once every 10,000 years (0.3% in 30 years)
Type of shore	No shores 	Shores next to dike 	Shores off the coast 
Bird population	Decrease of 30% 	No change 	Decrease of 10% 
Annual local tax	No change	60 Euro per year MORE	100 Euro per year MORE
You prefer:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

(b)

	OPTION 1 Increase in dike height	OPTION 2 No increase in dike height	OPTION 3 No increase in dike height
Flood probability	Once every 10,000 years (0.3% in 30 years)	Once every 2,000 years (1.5% in 30 years)	Once every 1,000 years (3% in 30 years)
Type of shore	Shores next to dike 	No shores 	Shores off the coast 
Bird population	No change 	Decrease of 30% 	Decrease of 10% 
Annual local tax	No change	60 Euro per year LESS	100 Euro per year LESS
You prefer:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Fig. 3** a. Example choice card experiment WTP2 b. Example choice card experiment WTA2. (Color figure online)

total number of respondents who completed the survey is 1,208 and is more or less equally divided across the 4 subsamples ( $n = 298, 297, 314, 299$ ), yielding an average overall response rate of around 80 %.

The socio-demographic respondent characteristics in the four samples and for the population from which the samples were drawn are presented in Table 3. The table clearly shows that the socio-demographic composition of the four samples is very similar to the composition of the overall population from which the samples were drawn. Also there are no significant differences between samples. Any differences observed in the estimation results between the four experiments are therefore not expected to be caused by differences between samples in socio-demographic background and household composition.

## 5 Statistical Model and Tests

Preferences are modelled in terms of McFadden's (McFadden 1974) Random Utility Model (RUM), allowing for a separation of utility ( $U_{ijt}$ ) into a deterministic ( $V_{ijt}$ ) and a stochastic part ( $\varepsilon_{ijt}$ ). Choice experiments fall in the class of attribute-based methods in which the deterministic part of utility of individual  $i$  for good  $j$  in choice task  $t$  is described in Eq. (4) as a linear function of its attributes  $X_{ijt}$  and other explanatory variables  $Z_{ijt}$ :

**Table 3** Socio-demographic characteristics for the four sub-samples and for the population from which they were drawn

	Population (%)	WTP1(%)	WTP2(%)	WTA1(%)	WTA2(%)
<i>Gender</i>					
Male	49.4	50.3	50.8	49.0	49.8
Female	50.6	49.7	49.2	51.0	50.2
<i>Age composition</i>					
Age < 25	11.0	10.1	9.4	9.6	10.7
25 => Age < 35	11.8	13.1	13.8	15.6	14.7
35 => Age < 45	19.3	20.1	16.8	17.8	18.7
45 => Age < 55	22.1	18.8	20.5	21.3	19.4
55 => Age < 65	20.4	20.5	17.5	16.6	20.4
Age >= 65	15.9	17.4	21.9	19.1	16.1
<i>Social class (education and profession)</i>					
1 (High)	21.2	21.8	20.9	23.2	23.4
2	34.3	34.2	31.6	36.0	34.4
3	15.2	14.8	14.1	13.4	15.7
4	26.3	27.2	29.0	25.2	23.4
5 (Low)	3.0	2.0	4.4	2.2	3.0
<i>Household size</i>					
1	17.6	18.1	20.2	17.2	18.7
2	35.1	37.9	37.4	38.9	37.1
3	17.6	17.1	15.8	18.8	16.7
4	20.2	19.8	20.2	18.5	18.1
> 4	9.5	7.0	6.4	6.7	9.4
<i>Education</i>					
1 (Low)	—	0.7	0	0	0.3
2	—	3.0	4.4	2.9	3.3
3	—	18.1	20.5	17.8	17.1
4	—	12.1	14.8	9.6	13.0
5	—	26.2	20.9	23.9	23.4
6	—	14.4	15.8	16.9	15.1
7	—	17.8	17.8	22.3	19.1
8 (high)	—	7.4	5.7	6.1	8.4
Unknown	—	0.3	0	0.6	0.3
<i>Number of respondents</i>	6,803	298	297	314	299

$$U_{ijt} = V_{ijt} + \varepsilon_{ijt} = \beta X_{ijt} + \alpha Z_{ijt} + \varepsilon_{ijt}, \quad \forall j \in K. \quad (4)$$

In each choice task  $t$  respondent  $i$  is presented with a limited set of possible choice alternatives  $K$ , in our case consisting of a reference scenario and two policy scenario's. The stochastic term is assumed to follow an IID extreme value distribution of type 1.

To account for preference heterogeneity the preference parameters for the attributes are allowed to vary across respondents. Equation (5) describes the mixed logit (ML) probability

$P$  of individual  $i$  selecting alternative  $j$  in choice task  $t$  from the set of available choice alternatives  $K$ :

$$P_{ijt} = \int \left( \frac{\exp(\beta_i X_{ijt} + \alpha Z_{ijt})}{\sum_{k=1}^K \exp(\beta_i X_{ikt} + \alpha Z_{ikt})} \right) \Delta(\beta_i | b) d\beta_i, \quad \forall j \in K. \quad (5)$$

The utility coefficients  $\beta$  for the non-price attributes vary across individuals (hence  $\beta_i$ ) with probability density  $\Delta(\beta_i | b)$ . This density can be a function of any set of parameters and represents in this case the mean and variance of  $\beta$  in the population sample.

Applications of ML-models have shown that this model is superior to the standard multinomial logit model in terms of overall fit and accuracy of welfare estimates (e.g., [Provencher and Bishop 2004](#); [Brouwer et al. 2010](#)). Mixed logit models account for preference heterogeneity and repeated choices ([Train 2003](#)). The price attribute included in the choice experiment implies that monetary welfare estimates can be derived (e.g., [Hensher et al. 2005](#)), representing the WTP or WTA arising from a change in the bundle of policy characteristics.

Our estimation results are used for hypothesis testing. We test hypothesis 1 by comparing the WTA and WTP for identical flood probability levels ( $H_0$ :  $WTA1 > WTP1$  &  $WTA2 > WTP2$ ). For this we use Poe's combinatorial test procedure ([Poe et al. 2005](#)). Hypothesis 2 is tested by comparing WTP (WTA) estimates for identical flood probability levels from experiments 1 and 2 ( $H_0$ :  $WTP2 > WTP1$  &  $WTA2 > WTA1$ ), while hypothesis 3 is tested by comparing the WTA–WTP disparity from experiment 1 with the WTA–WTP disparity from experiment 2 ( $H_0$ :  $WTA2/WTP2 > WTA1/WTP1$ ). For testing hypotheses 2 and 3, using Poe's combinatorial test procedure is not possible because the underlying statistical distributions for these transformed WTP and WTA values are unknown, so we rely on the point estimates for hypothesis testing in these two cases.

Before turning to the estimation results in the split samples, we furthermore performed statistical tests following [Swait and Louviere \(1993\)](#) to assess whether the estimated models for the four choice experiments are statistically significantly different. The [Swait and Louviere \(1993\)](#) test procedure is aimed at comparing equality of preference and scale parameters of the estimated choice models. Scale parameters and hence choice variance may differ across repeated choice sequences, for instance due to preference learning ([Brouwer et al. 2010](#)), but also because of possible differences between contexts in which the experiments were conducted, in this case the four versions of the choice experiment. Comparison of the estimated coefficients for the choice attributes is not straightforward due to the confounding of preference and scale parameters. This prevents the attribution of observed differences to both differences in scale and preferences or preference parameter inequality and equal scales ([Louviere et al. 2000](#)). The Likelihood Ratio (LR) test procedure therefore consists of a two-step approach where the scale parameter of one of the two datasets is first rescaled to allow testing of equality of preference parameters, whilst controlling for varying scale parameters between the two choice data sets. Secondly, a test for scale parameter equality is performed. The latter test is only meaningful if the preference parameters are found to be equal between the analysed choice sets. We run the test procedure for four combinations of the experiments, i.e., WTP1 versus WTP2, WTA1 versus WTA2, WTP1 versus WTA1, and WTP2 versus WTA2. The test results show that the first null hypothesis of equality of preference parameters is convincingly rejected for all four comparisons at the 1 % level. This implies that the estimated four choice models are not the same. However, as mentioned we are unable to attribute this to either the inequality of preferences or the scale parameters.

## 6 Estimation Results

Knowing that the four choice models are significantly different, we present the results of the mixed logit models separately for the WTP experiments in Table 4 and for the WTA experiments in Table 5. Tax is included as a continuous variable and dummy variables are included for all other attribute levels to test for potential non-linear effects. The reference categories for the dummy variables in the WTP and WTA models are presented in Table 2. The constant is included in the reference scenario utility function to test whether increasing dike height has an effect on safety perception, whilst controlling for changes in flood probabilities. The mixed logit models are estimated using 500 Halton draws from a uniform distribution for the dummy variables (as suggested by [Hensher et al. 2005](#)), in order to improve the efficiency of the statistical estimation procedure ([Bhat 2003](#)). Tax is included as a fixed parameter because this generally results in lower standard deviations of the estimated welfare measures ([Hensher et al. 2005](#)). WTP and WTA standard errors are obtained using the Delta method (see [Greene 2003](#)).<sup>4</sup>

The estimated standard deviations of the random parameter distributions in Table 4 show that, judging by their t-values, there is preference heterogeneity for increasing the height of dikes, reducing flood probabilities to once every 5,000 and 10,000 years, creating shores next to the dike and preserving the bird population. Estimated means of the parameters have the expected signs for all attribute levels and are statistically significant at a significance level of at least 5 %, except for the reference scenario constant in WTP1. The reference scenario constants in the models measure preferences for increasing the height of dikes. The large standard deviation of the parameter distribution on raising dikes shows that both positive and negative preferences are found in our samples. Negative preferences may be due to the fact that increasing dike height limits or takes away the view on the IJsselmeer at those locations where there currently is a view. Positive preferences may be motivated by an increase in safety perception. WTP for creating shores along the coast of the IJsselmeer is substantial. In both experiments people appear to prefer shores next to the dike to shores off the coast. However, the differences are small and Wald tests on the equality of the parameters reveal that the differences are statistically insignificant for both WTP1 and WTP2. WTP for (partly) preserving the bird population is substantial as well, with marginal WTP slightly decreasing when bird preservation levels increase.

WTP estimates for flood probabilities that are two and ten times smaller than in the reference scenario are very similar for experiments WTP1 and WTP2. Given the relatively small difference in WTP for a flood probability level of once every 10,000 (5,000) years and a flood probability level of once every 2,000 (1,000) years, the results show that there is a strong decreasing marginal WTP for reductions in flood probability. Poe tests reveal that the differences in WTP for experiment 1 are still statistically significant at <2 %, but for experiment 2 they are statistically insignificant at 10 %.

The patterns in the results from the WTA experiments (see Table 5) are very similar to those for the WTP experiments. We find substantial preference heterogeneity, increasing the height of dikes is valued positively, shores are preferred to no shores, (partly) preserving the bird populations is valued positively, and an increase in flood probability substantially

<sup>4</sup> Respondent characteristics for the four experiments are very similar (see Table 3), implying that it is unlikely that they explain the findings presented in this paper. In order to confirm this we estimated models with interactions between the choice attributes and three potentially important background statistics, i.e., income, distance to the IJsselmeer and safety perception. The results show that the effects are small and/or statistically insignificant, and that they do not affect the main patterns and findings. We therefore chose to present the attributes-only models in this section.

**Table 4** Mixed logit estimation results for the two WTP experiments

Attributes	WTP1			WTP2				
	Coefficient	t value	wtp	se(wtp)	Coefficient	t value	wtp	se(wtp)
Means of random parameters								
Reference scenario constant								
No increase in dike height	-1.517	-4.57	-€ 83	€ 18.3	-0.494	-1.40	-€ 25	€ 17.5
Flood probability								
1 in 500 → 1 in 1,000	0.693	8.37	€ 38	€ 4.5	-	-	-	-
1 in 500 → 1 in 5,000	0.826	8.14	€ 45	€ 5.4	-	-	-	-
1 in 1,000 → 1 in 2,000	-	-	-	-	0.681	7.51	€ 34	€ 4.5
1 in 1,000 → 1 in 10,000	-	-	-	-	0.993	8.55	€ 50	€ 5.6
Type of shore								
No → Off the coast	0.366	4.40	€ 20	€ 4.5	0.425	4.75	€ 21	€ 4.4
No → Next to dike	0.513	5.18	€ 28	€ 5.2	0.576	5.39	€ 29	€ 5.2
Bird population								
-30% → -10%	0.632	7.50	€ 35	€ 4.5	0.922	9.50	€ 46	€ 4.7
-30% → 0%	0.870	8.46	€ 48	€ 5.3	1.070	8.27	€ 54	€ 6.1
Standard deviations of random parameters								
Reference scenario constant								
No increase in dike height	6.844	13.7	-	-	7.007	13.3	-	-
Flood probability								
1 in 500 → 1 in 1,000	0.341	0.65	-	-	-	-	-	-
1 in 500 → 1 in 5,000	1.353	6.70	-	-	-	-	-	-
1 in 1,000 → 1 in 2,000	-	-	-	-	0.157	0.45	-	-
1 in 1,000 → 1 in 10,000	-	-	-	-	1.658	7.93	-	-

Table 4 continued

Attributes	WTP1			WTP2				
	Coefficient	t value	wtp	se(wtp)	Coefficient	t value	wtp	se(wtp)
<b>Type of shore</b>								
No → Off the coast	0.487	1.32	—	—	0.253	0.48	—	—
No → Next to dike	1.301	6.57	—	—	1.274	5.84	—	—
<b>Bird population</b>								
−30% → −10%	0.227	0.58	—	—	0.819	2.63	—	—
−30% → 0%	1.348	6.98	—	—	2.133	9.61	—	—
<i>Fixed parameters</i>								
<b>Annual local tax</b>								
Increase	−0.018	−18.2	—	—	−0.020	−18.0	—	—
<b>Model fit</b>								
Number of observations	2,980				2,970			
Log-Likelihood	−2,106				−2,034			
Pseudo R <sup>2</sup> (adjusted)	0.355				0.375			

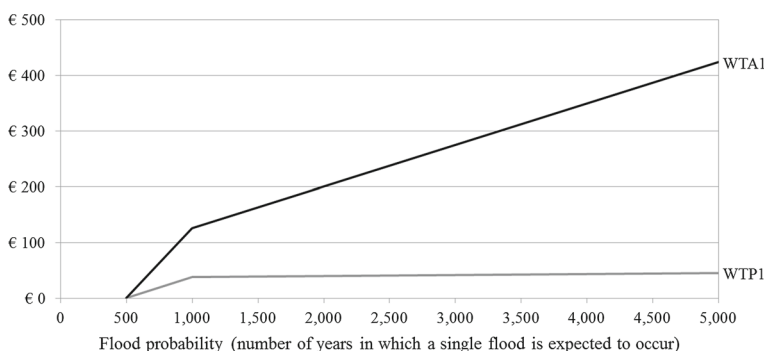


**Table 5** Mixed logit estimation results for the two WTA experiments

Attributes	WTA1			WTA2				
	Coefficient	t value	wta	se(wta)	Coefficient	t value	wta	se(wta)
Means of random parameters								
Reference scenario constant								
Increase in dike height	1.056	3.77	−€ 188	€ 51.0	1.073	3.31	−€ 158	€ 46.6
Flood probability								
1 in 5,000 → 1 in 1,000	−1.678	−7.37	€ 298	€ 62.3	−	−	−	−
1 in 5,000 → 1 in 500	−2.384	−8.57	€ 424	€ 81.0	−	−	−	−
1 in 10,000 → 1 in 2,000	−	−	−	−	−1.313	−7.43	€ 193	€ 36.7
1 in 10,000 → 1 in 1,000	−	−	−	−	−1.612	−9.09	€ 237	€ 39.9
Type of shore								
Next to dike → Off the coast	−0.301	−2.27	€ 53	€ 24.7	0.020	0.15	−€ 3	€ 19.5
Next to dike → No	−1.035	−5.27	€ 184	€ 41.2	−0.443	−3.08	€ 65	€ 21.8
Bird population								
0% → −10%	−0.955	−6.62	€ 170	€ 34.6	−1.120	−7.15	€ 165	€ 29.8
0% → −30%	−1.885	−8.74	€ 335	€ 62.4	−2.458	−9.76	€ 361	€ 58.0
Standard deviations of random parameters								
Reference scenario constant								
Increase in dike height	5.055	14.8	−	−	5.529	13.8	−	−
Flood probability								
1 in 5,000 → 1 in 1,000	2.622	6.30	−	−	−	−	−	−
1 in 5,000 → 1 in 500	3.175	7.78	−	−	−	−	−	−
1 in 10,000 → 1 in 2,000	−	−	−	−	1.759	5.01	−	−
1 in 10,000 → 1 in 1,000	−	−	−	−	1.758	5.23	−	−

Table 5 continued

Attributes	WTA1			WTA2				
	Coefficient	<i>t</i> value	wta	se(wta)	Coefficient	<i>t</i> value	wta	se(wta)
<b>Type of shore</b>								
Next to dike → Off the coast	1.252	3.57	—	—	1.449	4.62	—	—
Next to dike → No	2.342	7.68	—	—	1.199	3.52	—	—
<b>Bird population</b>								
0% → −10%	1.303	3.81	—	—	1.813	5.67	—	—
0% → −30%	2.662	7.92	—	—	3.004	8.11	—	—
<i>Fixed parameters</i>								
<b>Annual local tax</b>								
Decrease	0.006	5.79	—	—	0.007	7.04	—	—
<b>Model fit</b>								
Number of observations	3,140				2,990			
Log-Likelihood	−1,877				−1,766			
Pseudo R <sup>2</sup> (adjusted)	0.455				0.461			



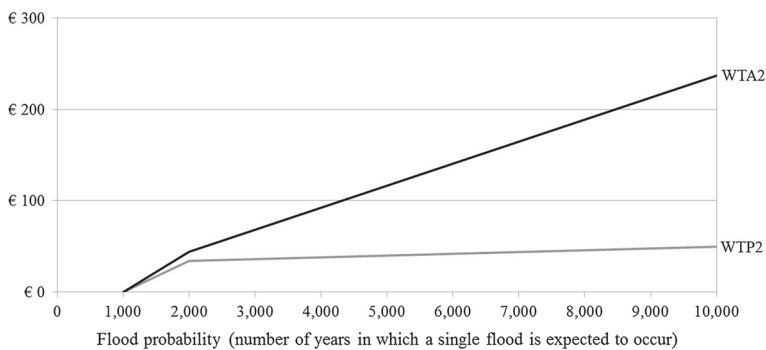
**Fig. 4** WTA/WTP estimates for avoiding/accepting an increase in flood probability to once every 500 years (in Euros per year) in experiments WTP1 and WTA1

reduces utility. The differences in WTA for different flood probability levels are larger than for the WTP experiments, i.e., marginal WTA is still decreasing, but less so than for WTP. Still, the Poe tests reveal that both for experiment 1 and 2 the WTA values for the different flood probability levels are statistically insignificant at 10 %. The most striking difference between the WTP and WTA models is that the estimated marginal WTA values are substantially higher than the estimated marginal WTP values. In the next section we analyse this issue in more detail.

## 7 Reference Dependence Effects

In order to test our first hypothesis and show the differences between WTA and WTP more clearly, we present in Fig. 4 the marginal WTP and WTA estimates for avoiding or accepting an increase in flood probability to once every 500 years for experiments WTP1 and WTA1. Both the WTP and WTA curve show a decrease in the slope of the flood probability curves, but where decreasing marginal WTA is small, decreasing marginal WTP is large. Performing a Wald test on the differences in slope parameters in Tables 4 and 5 reveals that they are statistically significant at 1 percent for WTP1, and statistically significant at only 10 % for WTA1. As a consequence the difference between WTP and WTA increases when the change away from the reference situation increases. Also WTA is almost 2.5 times larger than WTP for a flood probability of once every 1,000 years, and the difference is statistically significant at the 10 % level according to the Poe test. For a flood probability of once every 5,000 years, WTA is nearly eight times larger than WTP, and the Poe test results show that this difference is statistically significant at 1 %.

In Fig. 5 we present the same WTA–WTP estimates for avoiding or accepting an increase in flood probability compared to a reference point of once every 1,000 years for experiments WTP2 and WTA2. The patterns are similar to those in Fig. 4: the decrease in marginal WTA is substantially smaller than the decrease in marginal WTP, and the difference between WTP and WTA increases when moving away from the reference point. Performing a Wald test on the differences in flood probability slope parameters in Tables 4 and 5 reveals that the differences in the slopes of the value functions are statistically significant at the 1 % level for WTP2, but insignificant for WTA2. For a flood probability of once every 2,000 years the WTA is only somewhat larger than the WTP, and is statistically insignificant according to Poe's



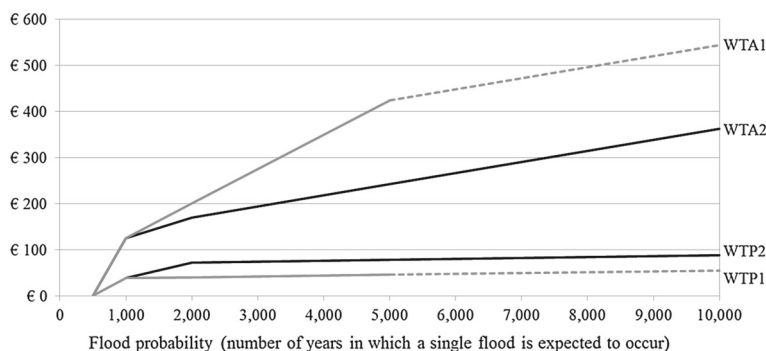
**Fig. 5** WTP/WTA estimates for avoiding/accepting an increase in flood probability to once every 1,000 years (in Euros per year) in experiments WTP2 and WTA2

combinatorial test. For a flood probability of once every 10,000 years WTA is nearly five times larger than WTP, and the Poe test shows that the difference is statistically significant at 1 %.

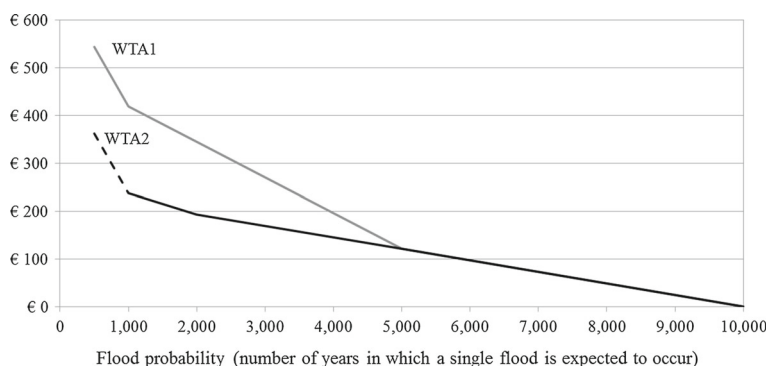
To address the second hypothesis, we compare the results from WTP1 with those from WTP2, and the results from WTA1 with those from WTA2, while the third hypothesis is addressed by comparing the WTA–WTP disparities obtained from experiments 1 and 2. In order to be able to better compare these results we give the WTP and WTA curves obtained from experiments 1 and 2 identical reference points. This can be done in two ways. For clarity's sake, we present and discuss the results from the approach where we change the reference point for experiment 2 to that of experiment 1 (i.e. a flood probability of once every 500 years), by augmenting the WTP and WTA curves from experiment 2 with the WTP and WTA values for once every 1,000 years from experiment 1. We furthermore extrapolate the WTP and WTA curves in experiment 1 by using the WTP and WTA slope parameters from experiment 2.<sup>5</sup>

The results are presented in Fig. 6. The figure reveals several things. First, the *curvature* of the two WTP lines from the two experiments are very comparable, but the *WTP values* for specific flood probabilities are larger for experiment 2 than for experiment 1. Using a lower flood probability reference point in experiment 2 leads to an upward shift in the value function. Without reference dependence and given an identical reference point, one would expect the WTP for a flood probability of once every 5,000 years to be at least as large as the WTP for a flood probability of once every 2,000 years. The reason is that people are expected to positively value a gain when comparing once every 5,000 years to once every 2,000 years, or at least be indifferent between the two flood probabilities. This is, however, not the case: WTP for a flood probability of once every 5,000 years is almost 30 Euros lower than for once every 2,000 years, implying a substantial impact of the reference point. This suggests a pattern resembling some form of reference point updating as discussed in Sect. 2, and implies that the model in, among others, [Viscusi and Huber \(2012\)](#) may have to be extended to incorporate such updating behaviour by including some form of weight function  $\lambda = f(p_n|p_r)$ .

<sup>5</sup> The second approach would be to use the slopes of the WTA and WTP curves from experiment 2 to extrapolate downward and obtain a change in flood probability from once every 1,000 years to once every 500 years for experiment 2. This approach is considered less preferable since we linearly extrapolate a clearly non-linear value function.



**Fig. 6** WTA and WTP value functions for accepting/avoiding an increase in the flood probability to once every 500 years (in Euros per year; dotted lines are extrapolations)



**Fig. 7** Alternative WTA value functions for accepting an increase in the flood probability to once every 500 years (in Euros per year; dotted line is an extrapolation)

Using a different reference point furthermore has a substantial effect on WTA, but using a lower flood probability reference point in experiment 2 now leads to a downward shift in the value function. At first sight it therefore appears that the WTA pattern is the exact opposite of the pattern for WTP. However, note that the property rights and the associated reference points for the WTA experiments are not the same as for the WTP experiments. For analysing possible reference point updating, the WTA value functions should therefore be drawn starting from their actual reference points, i.e., with a flood probability of once every 5,000 years for WTA1 and once every 10,000 years for WTA2. To make these two value functions comparable we again change the reference point for WTA1 to once every 10,000 years by augmenting it using the WTA2 value function, similar to the process used in Fig. 6. We furthermore extrapolate the WTA2 flood probability function from once in 1,000 years to once in 500 years using the WTA1 value estimate for that range. The resulting WTA value functions are shown in Fig. 7.

When analysing and interpreting the patterns observed in Fig. 7, it is important to realize that, for WTA, experiment 1 is the experiment with the updated reference point. Figure 7 clearly displays the form of reference point updating discussed in Section 2, i.e., an upward shift in the WTA value function due to a change in the reference point. For example, the WTA value for accepting an increase in flood probability to once every 1,000 years is around 1.8 times higher for WTA1 than for WTA2. Our findings therefore suggest similar forms of

reference point updating for WTA and WTP. Moreover, the effects appear to be larger for WTA than for WTP, at least in absolute terms, implying the weight function  $\lambda = f(p_n|p_r)$  is steeper for losses than for gains, i.e.,  $\lambda^- > \lambda^+$ .

## 8 Conclusions and Discussion

In this study we tested the effects of reference dependence in a low probability-high impact risk context. We assessed the effects of both gains and losses in flood probability, and of changing the flood probability reference value. Our prior expectations were that: (1) there is a disparity between WTP and WTA; (2) under risk aversion decreasing the flood probability reference value moves both the WTP and WTA value functions upward, for which we introduce the concept of reference point updating; (3) the WTA–WTP disparity increases as we increase the flood probability reference. To test these hypotheses we conducted four separate choice experiments using four split samples with similar socio-demographic characteristics living in the same case study area facing similar flood exposure levels, resulting in two WTP experiment with different flood probability reference values, and two WTA experiments with different flood probability reference values.

With respect to our first hypothesis, we find that WTA values are larger than WTP values for identical changes in flood probability. Moreover, we find that the disparity increases when changes away from the reference point are larger, a finding that is comparable to the results obtained by [Chilton et al. \(2012\)](#) in the context of health risks. With respect to our second hypothesis, we find that both the WTP and WTA value functions for changes in flood probabilities are affected by the reference point used in the choice experiment, and that the effects are stronger for losses than for gains. Specifically, when decreasing the flood probability reference value (lower flood probability) the WTP value function moves upward. The pattern for the WTA value function is identical, but because property rights and reference points were defined differently for WTA than for WTP, the value function moves upward when increasing the flood probability reference value (higher flood probability). As a result, with respect to our third hypothesis, we find that the WTA–WTP disparity increases when using a lower flood probability reference value, or vice versa the disparity decreases when using a higher reference value. These findings confirm the patterns expected under reference point updating.

Our findings have several implications, especially in the field of public good valuation. First, the effects of changing the reference value and reference point updating imply that due to different reference values between regions and between time periods there may be strong temporal and spatial dynamics in the welfare consequences of environmental changes. More specifically, welfare consequences of similar environmental changes may vary over time because reference values change gradually, and may differ between countries and regions because of location specific reference values. In order to reliably assess the welfare consequences of specific environmental changes, time- and space-specific research is therefore required.

A second and related point is that transferring values obtained from previous studies to other regions and time periods may lead to large transfer errors. Although this conclusion in itself is not new (see, e.g., [Dekker et al. 2011](#)), our findings suggest a potentially important reason why these errors are as large as they are. That is, if there are differences in the (risk) context and the reference (risk) levels between the situation in which existing values were elicited and the new situation to which these existing values are applied, not controlling

for these differences will lead to estimates that are off the mark. For meta-analysis value functions, which are often used for value transfer, this means that reference value effects need to be taken into account explicitly.

Finally, given the pervasive disparity between WTA and WTP, choosing the appropriate welfare measure to assess the economic consequences of changes in public good provision is essential. The use of WTP as the most widely endorsed indicator of welfare change may underestimate the economic value of welfare changes when it comes to accepting losses and may lead to suboptimal design of environmental policies (see also [Knetsch 2010](#)). However, our results show that WTA is more sensitive than WTP to both the scale of change that is being studied and the reference value. This implies that WTA values which are obtained from studies that assess a different range of possible changes, and that use a different reference value than is the case for the specific welfare analysis, may overestimate a welfare change. These observations call for case-study specific research on the welfare effects associated with changes in the quantity and/or quality of public environmental good provision. If value transfer is used instead, incorporating reference point effects in transferring values or functions is strongly advisable.

**Acknowledgements** The authors acknowledge funding from the PBL Netherlands Environmental Assessment Agency ([www.pbl.nl](http://www.pbl.nl)), from the ERA-Net BiodivERsA programme, with the national funder NWO, part of the 2011 BiodivERsA call for research proposals ([www.connect-biodiversa.eu/](http://www.connect-biodiversa.eu/)), and from the European Commission Seventh Framework Programme under Grant Agreement No. FP7-ENV-2012-308393-2 (OPERAs). We are thankful to Arjan Ruijs and Petra van Egmond (PBL) for cooperation on study design, to Celine Nobel and Nadja den Besten for pretest data collection, to Job van den Berg (TNS-NIPO) for assistance with the data collection, and to Pieter Jan Kerstens (VU University Amsterdam) for designing the photographs used in the survey. We also thank Wouter Botzen and two anonymous reviewers for very useful comments on earlier versions of the paper.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Bateman IJ, Day BH, Jones AP, Jude S (2009) Reducing gain–loss asymmetry: a virtual reality choice experiment valuing land use change. *J Environ Econ Manage* 58:106–118
- Bhat CR (2003) Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transp Res Part B* 37:837–855
- Bos F, Zwaneveld P, Van Puijenbroek PJTM (2012) Een Snelle Kosten-Effectiviteitanalyse voor het Deltaprogramma IJsselmeergebied (A Quick Cost-Effectiveness Analysis for the Deltaprogramma IJsselmeer Area). CPB Background Document, CPB Netherlands Bureau for Economic Policy Analysis, The Hague, The Netherlands
- Botzen WJW, Van den Bergh CJM (2009) Bounded rationality, climate risks and insurance: Is there a market for natural disasters? *Land Econ* 85:266–279
- Botzen WJW, Van den Bergh CJM (2012) Monetary valuation of insurance against flood risk under climate change. *Int Econ Rev* 53:1005–1025
- Brouwer R, Akter S, Brander L, Haque E (2009) Economic valuation of flood risk exposure and reduction in a severely flood prone developing country. *Environ Dev Econ* 14:397–417
- Brouwer R, Akter S (2010) Informing micro insurance contract design to mitigate climate change catastrophe risks using choice experiments. *Environ Hazards* 9:74–88
- Brouwer R, Dekker T, Rolfe J, Windle J (2010) Choice certainty and consistency in repeated choice experiments. *Environ Resour Econ* 46:93–109
- Brouwer R, Schaafsma M (2013) Modelling risk adaptation and mitigation behaviour under different climate change scenarios. *Clim Change* 117:11–29



- Brouwer R, Tinh BD, Tuan TH, Magnussen K, Navrud S (2014) Modeling demand for catastrophic flood risk insurance in coastal zones in Vietnam using choice experiments. *Environ Dev Econ* 19:228–249
- Chilton S, Jones-Lee M, McDonald R, Metcalf H (2012) Does the WTA/WTP ratio diminish as the severity of a health complaint is reduced? Wealth function. *J Risk Uncertain* 45:1–24
- Cummings RG, Brookshire DS, Schulze WD (1986) Valuing environmental goods. Rowman and Allanheld, Totowa, NJ, USA
- Dekker T, Brouwer R, Hofkes M, Moeltner K (2011) The effect of risk context on the statistical value of life: a Bayesian meta-model. *Environ Resour Econ* 49:597–624
- Dekker T, Koster P, Brouwer R (2014) Changing with the tide: semiparametric estimation of preference dynamics. *Land Econ* 90:717–745
- Greenberg E (2008) Introduction to Bayesian econometrics. Cambridge University Press, New York
- Greene WH (2003) Econometric analysis. Prentice Hall International, Inc., Upper Saddle River, New Jersey
- Hanemann WM (1991) Willingness to pay and willingness to accept: How much can they differ? *Am Econ Rev* 81:635–647
- Hensher DA, Rose JM, Greene WH (2005) Applied choice analysis: a primer. Cambridge University Press, Cambridge
- Horowitz JK, McConnell KE (2002) A review of WTA/WTP studies. *J Environ Econ Manage* 44:426–447
- Horowitz JK, McConnell KE (2003) Willingness to accept, pay and the income effect. *J Econ Behav Organ* 51:537–545
- Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47:263–291
- Kind J (2014) Economically efficient flood protection standards for the Netherlands. *J Flood Risk Manage* 7:103–117
- Knetsch JL (2010) Values of gains and losses: reference states and choice of measure. *Environ Resour Econ* 46:179–188
- Kőszegi B, Rabin M (2006) A model of reference-dependent preferences. *Q J Econ* 121:1133–1165
- List JA (2003) Does market experience eliminate market anomalies? *Q J Econ* 118:41–71
- Loomes G, Orr S, Sudgen R (2009) Taste uncertainty and status quo effects in consumer choice. *J Risk Uncertain* 39:113–135
- Louviere JJ, Hensher DA, Swait JD (2000) Stated choice methods: analysis and applications. Cambridge University Press, Cambridge, UK
- McFadden D (1974) Conditional logit analysis of qualitative choice behaviour. In: Zarembka P (ed) *Frontiers in econometrics*. Academic Press, New York, pp 105–142
- Neilson WS (2006) Axiomatic reference-dependence in behavior toward others and toward risk. *Econ Theor* 28:681–692
- Poe GL, Giraud KL, Loomis JB (2005) Computational methods for measuring the difference of empirical distributions. *Am J Agric Econ* 87:353–365
- Provencher B, Bishop RC (2004) Does accounting for preference heterogeneity improve the forecasting of a random utility model? *J Environ Econ Manage* 48:793–810
- Sudgen R (1999) Alternatives to the neoclassical theory of choice. In: Bateman JJ, Willis KG (eds) *Valuing environmental preferences: theory and practice of the contingent valuation method in the US, EU and developing countries*. Oxford University Press, Oxford, UK, pp 259–301
- Swait J, Louviere J (1993) The role of the scale parameter in the estimation and comparison of multinomial logit models. *J Mark Res* 30:305–314
- Train K (2003) Discrete choice methods with simulation. Cambridge University Press, Cambridge
- Tversky A, Kahneman D (1991) Loss aversion in riskless choice: a reference-dependent model. *Q J Econ* 106:1039–1061
- Tversky A, Kahneman D (1992) Advances in prospect theory: cumulative representation of uncertainty. *J Risk Uncertain* 5:297–323
- Viscusi WK, Huber J (2012) Reference-dependent valuations of risk: Why willingness-to-accept exceeds willingness-to-pay. *J Risk Uncertain* 44:19–44